WHAT IS CLAIMED IS:

1. A method of identifying a zinc finger domain that recognizes a target site on a DNA, the method comprising:

(a) providing cells containing a reporter construct, the construct comprising a reporter gene operably linked to a promoter, wherein the reporter gene is expressed above a given level when a transcription factor recognizes both a recruitment site and a target site of the promoter, but not when the transcription factor recognizes only the recruitment site of the promoter;

(b) providing a plurality of hybrid nucleic acids, each of which encodes a non-naturally occurring protein comprising (i) a transcription activation domain, (ii) a DNA binding domain that recognizes the recruitment site, and (iii) a test zinc finger domain, wherein the encoded amino acid sequence of the test zinc finger domain varies among the members of the plurality;

(c) contacting the plurality of hybrid nucleic acids with the cells under conditions that permit at least one of the plurality of nucleic acids to enter at least one of the cells;

(d) maintaining the cells under conditions permitting expression of the hybrid nucleic acids in the cells; and

(e) identifying a cell that contains a hybrid nucleic acid of (b) and that expresses the reporter gene above the given level as an indication that the cell contains a hybrid nucleic acid encoding a test zinc finger domain that recognizes the target site.

2. The method of claim 1, wherein the cells are eukaryotic cells.

3. The method of claim 2, wherein the cells are yeast cells.

4. The method of claim 3, wherein the cells are *Saccharomyces cerevisiae* cells.

5. The method of claim 1, wherein the reporter gene is a selectable marker.

6. The method of claim 5, wherein the selectable marker is selected from the group consisting of *URA3*, *HIS3*, *LEU2*, *ADE2,* and *TRP1*.

7. The method of claim 1, wherein the reporter gene is selected from the group consisting of *lacZ*, CAT, luciferase, GUS, and GFP.

8. The method of claim 1, wherein the DNA binding domain comprises a zinc finger domain.

9. The method of claim 8, wherein the DNA binding domain comprises two zinc finger domains.

10. The method of claim 9, wherein the DNA binding domain comprises three zinc finger domains.

11. The method of claim 1, further comprising the steps of (i) amplifying a source nucleic acid encoding the test zinc finger domain from genomic nucleic acid, a messenger RNA (mRNA) mixture, or a complementary DNA (cDNA) mixture, using an oligonucleotide primer that anneals to a sequence encoding a conserved domain boundary to produce an amplified fragment; and (ii) utilizing the amplified fragment to construct a hybrid nucleic acid for inclusion in the plurality of hybrid nucleic acids of step (b).

12. The method of claim 1, further comprising the steps of (i) identifying a candidate zinc finger domain amino acid sequence in a sequence database; (ii) providing a candidate nucleic acid encoding the candidate zinc finger domain amino acid sequence, and (iii) utilizing the candidate nucleic acid to construct a hybrid nucleic acid for inclusion in the plurality of hybrid nucleic acids of step (b).

13. The method of claim 5, wherein the selectable marker is an auxotrophy gene required for the synthesis of a metabolite; the genome of the cells lacks a functional copy of the auxotrophy gene; and, during step (d), the cells are maintained in a medium prepared without the metabolite.

14. The method of claim 1, wherein steps (a) to (f) are repeated to identify a second test zinc finger domain that recognizes a second target site.

15. The method of claim 14, further comprising constructing a nucleic acid encoding a polypeptide comprising the first test zinc finger domain and the second test zinc finger domain.

16. A method of identifying a zinc finger domain that recognizes a target site on a DNA, the method comprising:

(a) providing cells containing a reporter construct, the construct comprising a reporter gene operably linked to a promoter, wherein the reporter gene is expressed above a given level when a transcription factor recognizes both a recruitment site and a target site of the promoter, but not when the transcription factor recognizes only the recruitment site of the promoter;

(b) amplifying a plurality of nucleic acid sequences, each of which encodes a test zinc finger domain, using an oligonucleotide primer that anneals to a nucleic acid encoding a conserved domain boundary;

(c) joining each nucleic acid sequence of (b) to nucleic acid sequences encoding (i) a transcription activation domain, and (ii) a DNA binding domain that recognizes the recruitment site, to form a plurality of hybrid nucleic acids;

(d) contacting the plurality of hybrid nucleic acids of (c) with the cells of (a) under conditions that permit at least one of the plurality of hybrid nucleic acids to enter at least one of the cells;

(e) maintaining the cells under conditions permitting expression of the hybrid nucleic acids in the cells; and

(f) identifying a cell that contains a hybrid nucleic acid of (c) and that expresses the reporter gene above the given level, wherein the hybrid nucleic acid encodes a zinc finger domain that recognizes the target site on a DNA.

17. The method of claim 16, wherein the cells are yeast cells.

18. The method of claim 16, wherein the reporter gene is selected from the group consisting of *lacZ*, CAT, luciferase, GUS, and GFP.

19. The method of claim 16, wherein the DNA binding domain comprises a zinc finger domain.

20. The method of claim 19, wherein the DNA binding domain comprises two zinc finger domains.

21. A method of determining whether a test zinc finger domain recognizes a target site on a promoter, the method comprising:

(a) providing a reporter construct comprising a reporter gene operably linked to a promoter, wherein the reporter gene is expressed above a given level when a transcription factor recognizes both a recruitment site and a target site of the promoter, but not when the transcription factor recognizes only the recruitment site of the promoter;

(b) providing a hybrid nucleic acid that encodes a non-naturally occurring protein comprising (i) a transcription activation domain, (ii) a DNA binding domain that recognizes the recruitment site, and (iii) a test zinc finger domain;

(c) contacting the reporter construct with a cell under conditions that permit the reporter construct to enter the cell;

(d) prior to, after, or concurrent with step (c), contacting the hybrid nucleic acid with the cell under conditions that permit the hybrid nucleic acid to enter the cell;

(e) maintaining the cell under conditions permitting expression of the hybrid nucleic acid in the cell; and

(f) detecting reporter gene expression in the cell, wherein a level of reporter gene expression greater than the given level is an indication that the test zinc finger domain recognizes the target site.

22. The method of claim 21, further comprising the step of amplifying a nucleic acid encoding the test zinc finger domain from genomic DNA, an mRNA mixture or a cDNA mixture using an oligonucleotide primer that anneals to a sequence encoding a conserved domain boundary.

23. The method of claim 21, further comprising the steps of (i) identifying a candidate zinc finger domain amino acid sequence in a sequence database; (ii) providing a candidate nucleic acid encoding the candidate zinc finger domain amino acid sequence, and (iii) utilizing the candidate nucleic acid to construct a hybrid nucleic acid for inclusion in the plurality of hybrid nucleic acids of step (b).

Subsa4

24. A method of determining whether a test zinc finger domain recognizes a target site on a promoter, the method comprising:

(a) providing a first cell comprising a reporter construct comprising a reporter gene operably linked to a promoter, wherein the reporter gene is expressed above a given level when a transcription factor recognizes both a recruitment site and a target site of the promoter, but not when the transcription factor recognizes only the recruitment site of the promoter;

(b) providing a second cell comprising a hybrid nucleic acid that encodes a protein comprising (i) a transcription activation domain, (ii) a DNA binding domain that recognizes the recruitment binding site, and (iii) a test zinc finger domain;

(c) fusing the first and second cells to form a fused cell;

(d) maintaining the fused cell under conditions permitting expression of the hybrid nucleic acids in the cell; and

(e) detecting reporter gene expression in the fused cell, wherein a level of reporter gene expression greater than the given level is an indication that the test zinc finger domain recognizes the target site.

25. The method of claim 24 wherein the first and second cells are yeast cells of the opposite mating types.

26. A method of determining whether a test zinc finger domain recognizes a target site on a promoter, the method comprising:

(a) providing a plurality of reporter constructs, each construct comprising a reporter gene operably linked to a promoter, wherein the reporter gene is expressed above a given level when a transcription factor recognizes both a recruitment site and a target site of the promoter, but not when the transcription factor recognizes only the recruitment site of the promoter;

(b) providing a cell containing a hybrid nucleic acid, that encodes a non-naturally occurring protein comprising (i) a transcription activation domain, (ii) a DNA binding domain that recognizes the recruitment site, and (iii) a test zinc finger domain;

(c) contacting the plurality of reporter constructs with the cell under conditions that permit at least one of the plurality of reporter constructs to enter the cell;

(d) maintaining the cell under conditions permitting expression of the hybrid nucleic acid in the cell; and

(e) identifying a cell that contains a reporter gene of (a) and that expresses the reporter gene above the given level as an indication that the reporter construct in the cell comprises a target site recognized by the test zinc finger domain.

27. The method of claim 26, wherein the target binding site is between two and six nucleotides long.

28. The method of claim 27, wherein the plurality of reporter constructs comprises every possible combination of A, T, G, and C nucleotides at at least two positions of the target binding site.

29. The method of claim 28, wherein the plurality of reporter constructs comprises every possible combination of A, T, G, and C nucleotides at at least three positions of the target binding sites.

30. The method of claim 26, wherein steps (a) to (e) are repeated for a second test zinc finger domain to identify a second binding preference.

31. The method of claim 30, further comprising constructing a nucleic acid encoding a polypeptide comprising the first second test zinc finger domains.

32. A method of identifying a plurality of zinc finger domains, the method comprising:

carrying out the method of claim 1 to identify a first test zinc finger domain; and

carrying out the method of claim 1 again to identify a second test zinc finger domain that recognizes a target site different from the target site recognized by the first test zinc finger domain.

33. A method of generating a nucleic acid encoding a chimeric zinc finger protein, the method comprising:

carrying out the method of claim 32;

constructing a nucleic acid encoding a polypeptide comprising the first and second test zinc finger domains.

34. A method of identifying DNA sequences recognized by zinc finger domains, the method comprising:

carrying out the method of claim 24 to identify a first target site recognized by a first test zinc finger domain; and

carrying out the method of claim 24 again to identify a second target site recognized by a second test zinc finger domain.

35. A method of generating a nucleic acid encoding a chimeric zinc finger protein, the method comprising:

carrying out the method of claim 34;

constructing a nucleic acid encoding a polypeptide comprising the first and second test zinc finger domains.

36. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Cys-X-Ser-Asn-X_b-X-Arg-His-X_{3-5}-His (SEQ ID NO:68),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

37. A nucleic acid comprising a sequence encoding the polypeptide of claim 36.

38. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-His-X-Ser-Asn-X_b-X-Lys-His-X_{3-5}-His (SEQ ID NO:69),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

39. A nucleic acid comprising a sequence encoding the polypeptide of claim 38.

40. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Ser-X-Ser-Asn-X_b-X-Arg-His-X_{3-5}-His (SEQ ID NO:70),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

41. A nucleic acid comprising a sequence encoding the polypeptide of claim 40.

42. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Ser-Thr-X_b-X-Val-His-X_{3-5}-His (SEQ ID NO:71),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

43. A nucleic acid comprising a sequence encoding the polypeptide of claim 42.

44. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Val-X-Ser-X_c-X_b-X-Arg-His-X_{3-5}-His (SEQ ID NO:72),

wherein X_a is phenylalanine or tyrosine, X_b is a hydrophobic residue, and X_c is serine or threonine.

45. A nucleic acid comprising a sequence encoding the polypeptide of claim 44.

46. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Ser-His-X_b-X-Arg-His-X_{3-5}-His (SEQ ID NO:73),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

47. A nucleic acid comprising a sequence encoding the polypeptide of claim 46.

48. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Ser-Asn-X_b-X-Val-His-X_{3-5}-His (SEQ ID NO:74),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

49. A nucleic acid comprising a sequence encoding the polypeptide of claim 48.

50. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Ser-X_c-X_b-X-Arg-His-X_{3-5}-His (SEQ ID NO:75),

wherein X_a is phenylalanine or tyrosine, X_b is a hydrophobic residue, and X_c is serine or threonine.

51. A nucleic acid comprising a sequence encoding the polypeptide of claim 50.

52. A purified polypeptide comprising an amino acid sequence 60% identical to SEQ ID NO:65.

53. A nucleic acid comprising a sequence encoding the polypeptide of claim 52.

54. A purified polypeptide, comprising an amino acid sequence 60% identical to an amino acid sequence selected from the group consisting of: SEQ ID NO:29, 127, 129, 131, 133, and 135.

55. A nucleic acid, comprising a sequence encoding the polypeptide of claim 54.

56. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Ala-His-X_b-X-Arg-His-X_{3-5}-His (SEQ ID NO:150),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

57. A nucleic acid comprising a sequence encoding the polypeptide of claim 56.

58. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Phe-Asn-X_b-X-Arg-His-X_{3-5}-His (SEQ ID NO:151),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

59. A nucleic acid comprising a sequence encoding the polypeptide of claim 58.

60. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Ser-His-X_b-X-Thr-His-X_{3-5}-His (SEQ ID NO:152),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

61. A nucleic acid comprising a sequence encoding the polypeptide of claim 60.

62. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Ser-His-X_b-X-Val-His-X_{3-5}-His (SEQ ID NO:153),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

63. A nucleic acid comrising a sequence encoding the polypeptide of claim 62.

64. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Ser-Asn-X_b-X-Ile-His-X_{3-5}-His (SEQ ID NO:154),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

65. A nucleic acid comprising a sequence encoding the polypeptide of claim 64.

66. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Ser-Asn-X_b-X-Arg-His-X_{3-5}-His (SEQ ID NO:155),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

67. A nucleic acid comprising a sequence encoding the polypeptide of claim 66.

68. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Thr-His-X_b-X-Gln-His-X_{3-5}-His (SEQ ID NO:156),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

69. A nucleic acid comprising a sequence encoding the polypeptide of claim 68.

70. A purified polypeptide comprising the amino acid sequence:

Cys-X_{2-5}-Cys-X_3-X_a-X-Gln-X-Thr-His-X_b-X-Arg-His-X_{3-5}-His (SEQ ID NO:157),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

71. A nucleic acid comprising a sequence encoding the polypeptide of claim 70.

72. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Arg-X-Asp-Lys-X_b-X-Ile-His-X_{3-5}-His (SEQ ID NO:158),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

73. A nucleic acid comprising a sequence encoding the polypeptide of claim 72.

74. A purified polypeptide comprising the amino acid sequence:

X_a-X-Cys-X_{2-5}-Cys-X_3-X_a-X-Arg-X-Ser-Asn-X_b-X-Arg-His-X_{3-5}-His (SEQ ID NO:159),

wherein X_a is phenylalanine or tyrosine, and X_b is a hydrophobic residue.

75. A nucleic acid comprising a sequence encoding the polypeptide of claim 74.

76. A purified polypeptide comprising an amino acid sequence 60% identical to SEQ ID NO:141.

77. A nucleic acid comprising a sequence encoding the polypeptide of claim 76.

78. A purified polypeptide comprising an amino acid sequence 60% identical to SEQ ID NO:107.

79. A nucleic acid comprising a sequence encoding the polypeptide of claim 78.

80. A purified polypeptide comprising an amino acid sequence 60% identical to SEQ ID NO:137.

81. A nucleic acid comprising a sequence encoding the polypeptide of claim 80.

82. A purified polypeptide comprising an amino acid sequence 60% identical to SEQ ID NO:145.

83. A nucleic acid comprising a sequence encoding the polypeptide of claim 82.

84. A purified polypeptide comprising an amino acid sequence 60% identical to SEQ ID NO:149.

85. A nucleic acid comprising a sequence encoding the polypeptide of claim 84.